AD-A072 132 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1

A VARIATION ON SCHEFFE'S THEOREM, WITH APPLICATION TO NONPARAME--ETC(U)

MAY 79 R J SERFLING N00014-76-C-0608

UNCLASSIFIED FSU-STATISTICS-M502 NL

1 OF 1

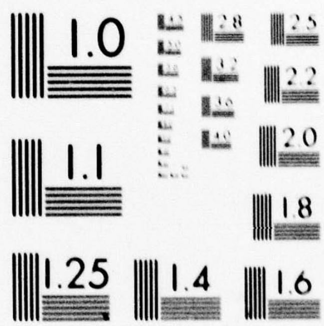
AD A072132

END

DATE FILMED

9-79

DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

12

6

A VARIATION ON SCHEFFE'S THEOREM, WITH APPLICATION TO NONPARAMETRIC DENSITY ESTIMATION

10

by R. J. Serfling

~~FSU Statistics Report M502~~
ONR Technical Report No. 137

9

11

May 1979

Department of Statistics
The Florida State University
Tallahassee, Florida 32306

12 11 P'

14

FSU-STATISTICS-M502, TR-137-ONR

15

Research supported by ~~the Army, Navy and Air Force~~ under Office of Naval Research Contract No. N00014-76-C-0608. Reproduction in whole or part is permitted for any purpose of the United States Government.

400 277

ABSTRACT

A VARIATION ON SCHEFFÉ'S THEOREM, WITH APPLICATION TO NONPARAMETRIC DENSITY ESTIMATION

For probability density functions $\{f_n\}$ and f defined on a d-dimensional Euclidean space, Scheffé (1947) proved the useful result that pointwise convergence of f_n to f implies convergence in the mean as well. In some applications it is of interest to know the rate of the mean convergence, and, in particular, to connect it with the rate of the uniform pointwise convergence. A basic lemma of this type is derived and applied to some problems in nonparametric density estimation.

Key phrases: Convergence in L_1-norm; Nonparametric density estimation.

1. INTRODUCTION

In real analysis and measure theory there are several results giving conditions under which pointwise convergence of a sequence of functions implies convergence in the mean as well. A specialized, but useful, such theorem was introduced into mathematical statistics by Scheffé (1947): if $\{f_n\}$ and f are probability density functions defined on d-dimensional Euclidean space R^d, and $\lim_n f_n(x) = f(x)$, $x \in R^d$, then $\lim_n \int |f_n(x) - f(x)| dx = 0$. In some applications it is of interest to know the rate of the mean convergence and, in particular, to connect it with the rate of the uniform pointwise convergence. A basic lemma of this type is derived in Section 2, and applications in nonparametric density estimation are considered in Section 3.

2. A BASIC LEMMA

For $x = (x_1, \ldots, x_d)$ in R^d, denote the usual Euclidean distance by $||x|| = (\sum_1^d x_i^2)^{\frac{1}{2}}$. For any function $h(x)$ on R^d, denote the sup-norm by $||x||_\infty = \sup_x |h(x)|$ and, for $p \geq 1$, the L_p-norm by $||h||_p = (\int |h|^p)^{1/p}$. For any probability density function f on R^d, define for each real $r > 0$ the (possibly infinite) constant

$$C_{f,r} = \sup_t t^r \int_{||x||>t} f(x)dx \quad (\leq E_f ||X||^r \leq \infty)$$

and for $r = \infty$ the constant $C_{f,\infty} = \sup \{t: \int_{||x||\leq t} f(x)dx = 1\}$. Also, denote by V_d the constant $\pi^{\frac{1}{2}d}\Gamma(\frac{1}{2}d + 1)$, which is the volume of the d-dimensional unit sphere. For two probability densities f and g, the following lemma provides a bound for $||g - f||_1$ in terms of $||g - f||_\infty$ and the constants just introduced.

LEMMA. *Let* f *and* g *be probability density functions on* R^d. *Then, for real* $r > 0$,

$$||g - f||_1 \leq 4V_d^{r/(r+d)} C_{f,r}^{d/(r+d)} \, ||g - f||_\infty^{r/(r+d)}, \tag{1}$$

and also

$$||g - f||_1 \leq 2V_d C_{f,\infty}^d ||g - f||_\infty. \tag{2}$$

PROOF. Put $h(x) = f(x) - g(x)$ if $f(x) \geq g(x)$ and $h(x) = 0$ otherwise. Then

$$||g - f||_1 = \int_{f \geq g}(f - g) - \int_{f<g}(f - g) = 2\int_{f \geq g}(f - g) - \int(f - g) = 2\int_{f \geq g}(f - g) = 2\int h.$$

Thus, for arbitrary constant $A > 0$ we have

$$||g - f||_1 = 2\int_{||x|| \leq A} h + 2\int_{||x|| > A} h \leq 2V_d A^d ||h||_\infty + 2A^{-r} C_{f,r}.$$

The two rightmost terms become equal for $A^{r+d} = V_d ||h||_\infty)^{-1} C_{f,r}$, in which case the righthand side becomes that of (1). A similar argument leads to (2). □

Clearly we may replace $C_{f,\alpha}$ by $\min \{C_{f,\alpha}, C_{g,\alpha}\}$ in the lemma, to produce a result symmetric in f and g. However, in typical applications the one-sided version is more apropos, as in the following result.

COROLLARY. *Let* f *be a probability density function on* R^d *satisfying* $C_{f,r} < \infty$ (*implied by* $E_f ||X||^r < \infty$), *where* $0 < r < \infty$. Let f_n *be a sequence of density functions satisfying* $\lim_n ||f_n - f||_\infty = 0$. *Then*

$||f_n - f||_1 = O(||f_n - f||_\infty^{r/(r+d)})$, $n \to \infty$. If, further, f has bounded support, then $||f_n - f||_1 = O(||f_n - f||_\infty$.

Thus, by strengthening the hypothesis of Scheffé's theorem, we obtain a rate for the convergence of $||f_n - f||_1$ to 0 in terms of that of $||f_n - f||_\infty$. Note that the property that we are dealing with probability density functions is crucial to the proof of the above lemma. Thus a generalization of the above corollary to arbitrary sequences $\{f_n\}$ converging pointwise to arbitrary f is not envisioned.

Along with the distances $||g - f||_\infty$ and $||g - f||_p$, let us also consider the Hellinger distance, $||g - f||_H = ||g^{\frac{1}{2}} - f^{\frac{1}{2}}||_2$. We then have the inequalities

$$||g - f||_H^2 \le 2||g - f||_1 \qquad (3)$$

and, for $p > 1$,

$$||g - f||_p \le ||g - f||_\infty^{p-1}||g - f||_1. \qquad (4)$$

These bounds on $||g - f||_\infty$ lead via the above Lemma to bounds on $||g - f||_p$ and on $||g - f||_H$.

3. APPLICATIONS IN NONPARAMETRIC DENSITY ESTIMATION

We now let f_n denote an empirical probability density function based on a sample of size n from a probability density function f defined on the real line. A broad literature offers many varieties of such f_n and many properties. In particular, under conditions slightly milder than f having a bounded first derivative, it has been shown for suitably designed estimators f_n that

$$||f_n - f||_\infty =_{wp1} O(n^{-1/3}\log n),\ n \to \infty. \tag{5}$$

(The exponent 1/3 in such results is subject to improvement only under additional regularity conditions on f, such as existence of a bounded second derivative.) In some situations it is of interest to establish the auxiliary convergence

$$||f_n - f||_2^2 = o_p\ (n^{-1/2}),\ n \to \infty. \tag{6}$$

We can obtain this result by using the Corollary in Section 2 in conjunction with (4), (5) and the additional assumption that $C_{f,r} < \infty$ for some $r > 1$ (implied by f having finite r-th moment for some $r > 1$).

As an example of when (6) is of interest, consider estimation of the quantity $\int f^2(x)dx$, which appears as a parameter in the Pitman asymptotic relative efficiencies of certain statistical procedures. When little is known about f, it may be of interest to estimate this parameter as a preliminary to selection of a statistical procedure. For the sample analogue estimator, $\int f_n^2(x)dx$, the corresponding asymptotic distribution theory may be derived as follows. Introduce the functional $T(f) = \int f^2(x)dx$ defined on the space of probability distributions f and approximate the estimation error $T(f_n) - T(f)$ by the first term in a Taylor expansion of T(f) about $T(f_n)$, following the technique of von Mises (1947). This approximation is found to be $2\int f(f_n - f)$, which for suitable type of f_n can be shown to be asymptotically normal by the use of central limit theory for double arrays. To transfer this asymptotic normality to the normalized estimation error $n^{1/2}[T(f_n) - T(f)]$, it is necessary to show that $n^{1/2}[T(f_n) - T(f) - 2\int f(f_n - f)] \to_p 0$. This reduces to showing

that (6) holds. (A more general treatment for a class of such efficacy-related parameters is in preparation with the collaboration of Kuang-Fu Cheng.)

Let us next consider the "mean integrated squared error," $E\{||f_n - f||_2^2\}$, for which as an analogue to (5) it is desirable to characterize the rate of convergence to 0. We confine attention to the case that f_n has the form $f_n(x) = n^{-1}a_n\sum_{i=1}^n K(a_n(x - X_i))$, $-\infty < x < \infty$, where $K(\cdot)$ is a known probability density function and $\{a_n\}$ is a sequence of positive constants tending to ∞. From Nadaraya (1974) we have

$$E\{||f_n - f||_2^2\} = O(a_n n^{-1}) + O(a_n^{-2s}), n \to \infty, \tag{7}$$

under some regularity conditions on K and assuming that the s-th derivative of f is a bounded $L_2(-\infty, \infty)$ function, with $s \geq 2$. In (7) the first term is the contribution due to the integrated variance, $\int E\{[f_n(x) - Ef_n(x)]^2\}dx$, the second term due to the integrated squared bias. We shall now handle the latter contribution by a different approach, using the above Lemma, and thus obtain a competitor to (7). For each $-\infty < z < \infty$ and $n = 1, 2, \ldots$, define the probability density function $g_{n,z}(x) = f(x - a_n^{-1}z)$, $-\infty < x < \infty$. Then $Ef_n(x) = \int g_{n,z}(x)K(z)dz$ and thus the integrated squared bias satisfies

$$\begin{aligned}\int[Ef_n(x) - f(x)]^2dx &= \int[\int(g_{n,z}(x) - f(x))K(z)dz]^2dx \\ &\leq \int\int[g_{n,z}(x) - f(x)]^2K(z)dz\, dx \\ &\leq \int\{\int[g_{n,z}(x) - f(x)]^2dx\}K(z)dz \\ &\leq M_{f,r}\int||g_{n,z} - f||_\infty^{(2r+1)/(r+1)}K(z)dz,\end{aligned}$$

where $M_{f,r}$ is a constant depending only on f and r. Here we have used

Jensen's inequality, Fubini's theorem, and the Lemma, along with the assumption that $C_{f,r}$ is finite (implied by the assumption that f has finite r-th moment.) If we further assume that f is Lipschitz on the real line, so that $||g_{n,z} - f||_\infty \le A_f a_n^{-1}|z|$, then we arrive at

$$E\{||f_n - f||_2^2\} = O(a_n n^{-1}) + O(a_n^{-(2r+1)/(r+1)}), \ n \to \infty. \tag{8}$$

Here we have assumed $\int |z|^2 K(z)dz < \infty$, as does Nadaraya (1974). With regard to f, our smoothness conditions are less stringent than Nadaraya's, but we have imposed a moment restriction. (A more general treatment allowing recursive versions of f_n is in preparation with the collaboration of Philip Cheng.)

As a final illustration, let us consider the mimimum Hellinger distance approach toward parametric robust efficient estimation, introduced by Beran (1977). In this theory a basic role is played by convergence in the Hellinger metric of a sequence of probability densities f_n to a limit f. By our Lemma and relation (3), we may derive rates for such convergence from given rates for convergence in the sup-metric. The rates for the Hellinger metric convergence have applications in connection with estimators represented as functionals, analogous to the treatment of the functional $T(f) = \int f^2$ discussed earlier.

This research was supported by the Statistics and Probability Program, Office of Naval Research, U.S. Department of the Navy, under Contract No. N00014-76-C-0608.

REFERENCES

BERAN, R. J. (1977). Minimum Hellinger distance estimates for parametric models. Ann. Statist. *5*, 445-463.

NADARAYA, E. A. (1974). On the integral mean square error of some nonparametric estimates for the density function. Th. Prob. Applic. *19*, 133-141.

SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. Ann. Math. Statist. *18*, 434-438.

VON MISES, R. (1947). On the asymptotic distributions of differentiable statistical functions. Ann. Math. Statist. *18*, 309-348.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M502 ONR No. 137	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE A VARIATION ON SCHEFFÉ'S THEOREM, WITH APPLICATION TO NONPARAMETRIC DENSITY ESTIMATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORGANIZATION REPORT NUMBER FSU Statistics Report M502
7. AUTHOR(s) R. J. Serfling		8. CONTRACT OR GRANT NUMBER(s) ONR No. N00014-76-C-0608
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics & Probability Program Arlington, Virginia 22217		12. REPORT DATE May, 1979 13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Convergence in L_1-norm; Nonparametric density estimation.

20. ABSTRACT

sub n

For probability density functions $\{f_n\}$ and f defined on a d-dimensional Euclidean space, Scheffé ~~(1947)~~ proved the useful result that pointwise convergence of f_n – sub n to f implies convergence in the mean as well. In some applications it is of interest to know the rate of the mean convergence, and, in particular, to connect it with the rate of the uniform pointwise convergence. A basic lemma of this type is derived and applied to some problems in nonparametric density estimation.